Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 99 (2016) 36 – 49
Procedia Computer Science

International Conference on Knowledge Management, ICKM 2016, 10-11 October 2016, Vienna, Austria

Using knowledge management to give context to analytics and big data and reduce strategic risk

John S. Edwards[a]*, Eduardo Rodriguez Taborda[b] *

[a]*Aston University, Operations and Information Management Group, Aston Triangle, Birmingham, B4 7ET, U.K.*
[b]*IQAnalytics Inc. Canada and Adjunct Professor Telfer School of Management University of Ottawa*

Abstract

At the moment, the phrases "big data" and "analytics" are often being used as if they were magic incantations that will solve all an organization's problems at a stroke. The reality is that data on its own, even with the application of analytics, will not solve any problems. The resources that analytics and big data can consume represent a significant strategic risk if applied ineffectively. Any analysis of data needs to be guided, and to lead to action. So while analytics may lead to knowledge and intelligence (in the military sense of that term), it also needs the input of knowledge and intelligence (in the human sense of that term). And somebody then has to do something new or different as a result of the new insights, or it won't have been done to any purpose.
Using an analytics example concerning accounts payable in the public sector in Canada, this paper reviews thinking from the domains of analytics, risk management and knowledge management, to show some of the pitfalls, and to present a holistic picture of how knowledge management might help tackle the challenges of big data and analytics.

Peer-review under responsibility of the Organizing Committee of ICKM 2016

Keywords: knowledge management; analytics; big data; data envelopment analysis; Canada; processes, strategic risk;

1. Introduction

Big data and analytics are currently two of the hottest topics in business and management literature and practice. Managers are bombarded with exhortations to learn about them and use them, and especially to persuade their

* Corresponding author. Tel.: +44-121-204-3236; fax: +44-121-204-5271.
E-mail address: j.s.edwards@aston.ac.uk

Peer-review under responsibility of the Organizing Committee of ICKM 2016
doi:10.1016/j.procs.2016.09.099

company to invest in software and hardware to help them to do it. Space does not permit us to include a full discussion here of the hype surrounding these phenomena at present. On the one hand, it has all the hallmarks of a "consultancy fad", with its tendency to be driven by IT; nevertheless, developments such as those in web analytics and real-time data do represent genuine opportunities for organizations.

We thus approach the development of analytics and big data in strategic risk management from two perspectives: strategic risk mitigant and strategic risk creator. Strategic risk can take different forms such as: reputation, industry margin reductions, economic efficiency, environmental issues, energy issues, geographical hazards, threats and terrorism, citizen-customer dissatisfaction, distribution of justice, and so on [1, 2]. In particular strategic risk in this context is the misuse of strategic resources such as data and analytics capabilities, and reduced organization/preparedness for using analytics and data in different formats on various types of data. This misuse of resources can be converted into a competitive position change or strategic risk [3]. This limited understanding of what analytics and big data are can reduce the chance of improving organizational value because of unsystematic use of data, tools, methods and the synergetic process between humans, machines and data. In addition, the investment of time and resources in analytics and big data represents a very significant opportunity loss if not used effectively.

What is evident is that so many of these exhortations discuss big data and analytics in isolation, lacking any form of context. Yet context is central to any management activity. A decision is only effective in its context - sometimes it will be appropriate to increase prices; at other times, to decrease them. Data on its own, no matter how big, even with the application of analytics, will not solve any problems. Any analysis of data needs to be guided, the context being one of the factors influencing that guidance, and to lead to appropriate action. Thus it needs the input of knowledge and intelligence. Analytics can identify useful patterns and relationships, but they must be based on solid and reliable foundations - GIGO (Garbage In, Garbage Out) is as true now as it ever was.

Our example in this paper concerns accounts payable. The activities of any organization can be divided into core and support processes [4]. The core processes are what the organization does for its external customers; the processes that literally make the organization what it is, the ones that distinguish an automobile manufacturer from a university or an insurance company. Support processes, as the name implies, support the core processes and are done for internal customers. Accounts payable is a classic example of a support process: every organization has to make sure it pays its bills. Even though it is not part of the organization's purpose and goals, it can still detract from the achievement of those goals, if not done well. Indeed, if done really badly, it could leave the organization unable to operate. A poor accounts payable process could be a very expensive process and thus a strategic issue to tackle for organizations. Such a process needs to be improved through a more effective approach that adds value to organizations. It can be part of implementing a more systematic approach to supplier relationship management in order to control other risks associated with supply chain and production/operations costs.

This example demonstrates that support processes too can generate big data, and does some preliminary analysis on this data using Data Envelopment Analysis, an analytics technique specifically designed to compare the performance of different organizations or operating units. We can safely assume that efficiency is the goal of accounts payable, as an accounts payable department has no choice about the things it does - it cannot choose to move into a new line of business!

We shall offer preliminary conclusions about the example, based on this pilot study, and also demonstrate the relevance of knowledge and knowledge management to steering the analytics process.

The paper is structured as follows. First, we briefly review the history of analytics and big data. We frame the issue of using analytics and big data as both mitigator and creator of strategic risk .Next we give more detail about data envelopment analysis, then explain the approach taken in our study. Following that, the example and its results so far are presented and we end with discussion and conclusions.

2. Analytics and big data

We take slightly unusual positions concerning both analytics and big data. Analytics we regard as a new name for a much older activity, whilst we believe that the "big" in big data is a relative term.

2.1. Analytics

Davenport and Harris[5] define analytics as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions" (p. 7). A commonly-found classification divides analytics techniques into three types: descriptive, predictive and prescriptive. There seem to be no agreed precise definitions of these, but they can be characterized as follows.

Descriptive analytics uses, for example, business intelligence, data mining, sentiment and affect analysis, web analytics, graph mining, to provide the context and trending information on past or current events, answering what has happened and what is happening.

Predictive analytics uses, for example, statistical models, machine learning, neural network analysis, and forecasts to provide an accurate projection of future happenings and the reasoning as to why, answering what could happen.

Prescriptive analytics uses, for example, optimization, simulation, artificial intelligence, case-based reasoning, to recommend one or more courses of action and shows the likely outcome of each decision, providing answers to what the organization should do.

Looking at the above definition and these more detailed examples, analytics can be seen to be virtually identical to the discipline of operations research/management science, which has been practiced under those names since around 1940. The professional society INFORMS, for example, gives the definition[6] "Operations Research (O.R.), or operational research in the U.K, is a discipline that deals with the application of advanced analytical methods to help make better decisions…Employing techniques from other mathematical sciences, such as mathematical modeling, statistical analysis, and mathematical optimization, operations research arrives at optimal or near-optimal solutions to complex decision-making problems."

So, analytics is a relatively new name for an activity with a long history. What has changed since 1940 is the ever-increasing range of techniques, most recently for dealing with textual and other non-quantitative data (see the following sections), and the nature of the IT support available to help with the process. There is a clear association of the work with two components for analytics practice: on the one hand the development of statistical techniques, on the other hand machine learning and computational capabilities. These two components complement the mathematical approach that has existed since the beginning of analytics, and provide more options to create knowledge from data.

2.2. Big data

Discussions of big data refer to it as being characterized by ever-growing values of the "three Vs", volume, variety and velocity [7, 8], with a fourth more stable V, veracity, often added to this list [9]. In the light of the ever-increasing capabilities of information technology, we feel that these can only be relative terms. What was a large volume of data in 2006 perhaps does not seem so large now. But that was large by 1996 standards, and so on. Definitions of big data therefore continue to evolve [10]. Looking at the way in which the big data debate is conducted, we would like to offer a rule of thumb: it's big data if you can't process it on your current desktop device.

An additional feature can be the incorporation of unstructured data [9, 11]. Managing this data requires data architectures that can deal with many types of data, such as graphical databases, rather than traditional data warehouses or data marts. Connecting multiple data sources enables new approaches to problems. For example, some years ago only a parametric approach to solving certain risk management problems was possible; these days the non-parametric approach brings a lot of advantages. Non-traditional means also now support decisions in financial transactions, such as using a process of mining discussions on social media related to companies' outlook.

2.3. Two illustrations

Analytics thus requires not only data and computational capacity, but also mathematical modeling and thinking in a systematic way. We will explain this with two small examples.

Consider first the following real-life example. A company running a chain of cafés wanted to know more about exactly what food they were selling/using, to avoid wasting fresh produce. So they turned to leading-edge information technology to collect and analyze data on a much greater scale than had been done previously. This sounds very similar to examples of big data and analytics that can be found in computing magazines almost every week at the moment.

But this was 1952! The company was J. Lyons and Co. Ltd. in the UK. Lyons was the first business to use computers - they had to manufacture their own, the LEO (Lyons Electronic Office) to do so - and this was one of their first applications.

They used what would nowadays be called (predictive) analytics, even taking into account the weather forecast, and for the time it was big data: the desktop devices of 1952 - manually-operated calculators - could not have produced the results in the time available (overnight, for the next day).

The second example [12] shows analytics as part of a deductive process and not just an inductive approach. “By 1941-42, the allies knew that US and even British tanks had been technically superior to German Panzer tanks in combat, but they were worried about the capabilities of the new marks IV and V.”[12] Observations by the army intelligence group produced contradictory answers about tank production, so they turned to analytics. The allies developed a formula, based on the serial numbers of captured tanks, and thus predicted that the Germans could produce around 246 tanks per month. “At that time, standard army intelligence estimates had believed the number was far, far higher, at around 1,400.”[12] After the war the records demonstrated that the actual number was 245 per month.

We trust we have made a convincing case that the activities in analytics are not completely new, and crucially therefore that there are lessons that can be learned from the experience of the past 75 years or more.

2.4. Linking analytics to business decisions

No-one disputes that analytics and big data are all about understanding - making sense of - the target phenomena. However, there are different types of understanding, and it isn’t possible to substitute one for another. Spender[13] has identified three types of this understanding, or organizational knowing, as he calls it: data, meaning and skilled practice. He also comments[14] that “The first two can be captured in language, the essence of practice cannot” (p.12) and that in necessarily embracing all three, knowledge management moves into areas that are “undetermined and inconclusive” (p.12). We would marginally disagree with Spender, as we believe that all knowledge has both explicit (codifiable in language, in Spender’s terms) and tacit (not readily expressible) components [15]. For data, the tacit element is very small indeed, as long as the assumptions made when collecting the data have also been written down and thus made explicit, not just the actual data. For meaning, the tacit element is small, but can still be significant - some data sets are easier to “read” than others. While for skilled practice, the tacit knowledge dominates, and it is the explicit element that is very small indeed (cf. riding a bicycle, where the explicit element does not go much further than “sit on the saddle and hold the handlebars”). In the context of analytics, both the meaning and the skilled practice span *two* domains - the target domain and the analytics domain itself. Also, meaning and skilled practice each have a reciprocal relationship with knowledge/knowing: they both influence it and are influenced by it. In an analytics study, the findings may lead to knowledge and intelligence (in the military sense of that term), but they also need the input of knowledge and intelligence (in the human sense of that term).

So, any application of analytics in business must be guided by the meaning and skilled practice from the business domain. The people running the business need to ask “What questions would we like answered that we can’t answer now?” or more generally, “What questions would we like answered better that we can’t answer well enough now?” As McAfee and Brynjolfsson[7] put it, “The first question a data-driven organization asks itself is not ‘What do we think?’ but ‘What do we know?’” (p.68).

Fig. 1 shows how analytics, (big) data and human knowledge fit together for a single study, starting from this guidance and leading to some purposeful action that makes a difference in the organization.

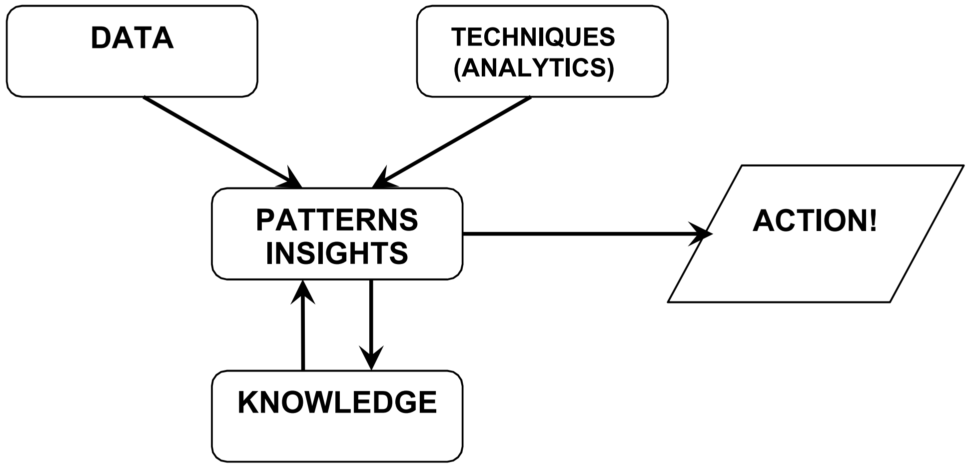

Fig. 1. The interactions between data, analytics and human knowledge in a single study.

2.5. Linking analytics to organizational risk

Organizations are moving from considering data just as something they have to deal with, at best as a tactical resource, to considering data as a means of creating value. This means managing data with a more strategic view. Additionally, the analytics process has the potential to be part of the improvement and development of the business's processes and strategy.

From the strategic risk management point of view, analytics is a bridge to reduce uncertainty and to move decisions to a landscape with better knowledge. As we know many questions cannot be answered precisely but, for example, the incorporation of randomness in what previously has only been analyzed as deterministic, can add value. If the randomness can be described by analytics models, the application can be developed to similar situations. In the example of accounts payable we probably can never predict what number of invoices that organization needs to process and how much the dollar value of these invoices will be. But it is possible to predict with some accuracy the range in which the values can fall and therefore help with budget definition, the number of employees assigned, the groups designed to control the process and so on.

Good organizational learning from the process of predictive analytics, or as in the example below through optimization approaches, can improve the process to reduce the risk of wasting money or the risk of lack of controls by segments of suppliers or groups of transactions.

The next section describes the analytics technique we will be using in the example in this paper.

3. Data envelopment analysis

Data Envelopment Analysis (DEA) is a linear programming procedure that is mainly used for finding the importance of inputs in generating outputs [16]. This is an approach to measuring relative efficiency on a scale from 0-100%. DEA assigns a score of 100% efficiency to a Decision Making Unit (DMU) only when comparisons with other DMUs do not provide evidence of inefficiency in the use of any input in the production of the output(s). That is, the mix of input-output for that DMU cannot be improved by using the "recipe" from one or more of the other DMUs. It is essential to bear in mind that DEA measures *relative* efficiency, i.e. efficiency compared to the set of DMUs analyzed. It is therefore often used for measuring the relative performance of similar DMUs within the same overall organization, such as branches of a bank or schools within the same district.

A DMU with an efficiency score less than 100% can be considered (relatively) inefficient. A score of less than 100% means that a linear combination of other DMUs from the sample could produce the same vector of outputs using a smaller vector of inputs. Graphical displays of DEA results often include the "production frontier" or "efficient frontier", which joins all combinations with 100% efficiency, though this can be difficult to show when there are more than two inputs or outputs. The efficiency score reflects the radial distance from the estimated production frontier to the DMU under consideration. Those DMUs with a score of 100% are located on the efficient frontier, while those that are comparatively less efficient are located below (nearer to the origin than) the efficient frontier.

The CCR Model [16] is the basic model used in DEA. The process generates an envelopment surface that can be represented using either a constant returns to scale (CRS) or a variable returns to scale (VRS) production system. In the CRS, it is assumed that outputs change in the same proportion as inputs do. Alternatively, VRS can be represented using increasing, constant and decreasing returns to scale. In this article we use CRS.

4. Study methodology

The study is based on data collected from six government organizations in Canada. They will be referred to by letters A-F in order to preserve confidentiality. For the same reason, numerical results presented here have all been multiplied by a constant. As DEA assumes linearity, this has no effect on measuring the relative performance of the DMUs, and in particular the efficiency scores. The theory outlined in section 2.4 and Fig. 1 will be used as a lens to understand and reflect on the modelling that has taken place, or that could take place now or on the future. One of our aims is to contrast these reflections with the misleading "just click the mouse and sit back" view of analytics often presented in the magazine literature.

5. Example

We will describe as much background as we can without breaking confidentiality, and then go on to explain which DEA investigations might be carried out, and why some are more useful than others.

5.1. Background

As mentioned earlier, all organizations must have an accounts payable process. It is believed that there is potential to improve the process of Accounts Payable (A/P) in Government of Canada organizations. This is the initial knowledge-based guidance, so the questions that the public sector wants answered are "can we run the A/P process better?" and if the answer to that is yes, "how can we run the A/P process better?" The six organizations that we studied have varying A/P processes; some are centralized, some decentralized, and some split. There is a blend of human resource used in the processes as well as a mix of activities to deal with. In the past, these organizations' A/P performance has been measured by just using ratios at an individual factor level. There was no attempt to consider a combination of factors, nor to carry out comparisons to the best performers by making use of all the factors involved in output production.

The challenge with this type of benchmarking study is that each organization likely operates their A/P process in many different ways. A traditional benchmarking approach, in which one might compare only ratios of inputs of a work unit (number of FTEs for example) to its outputs (services delivered), can therefore lead to an "apples to oranges" comparison when used to compare organizations. In contrast, the use of DEA allows for an input-output comparison across many DMUs while at the same time reviewing how the within-organization factors contribute to efficiency. This approach does not attempt to compare DMUs in a direct "head to head" manner. Rather, it compiles information from a number of DMUs to profile the "most efficient producer". All DMUs are then compared to this profile to determine their relative performance. The methodology used is applicable to any process and development of standards across organizations to improve productivity.

The type of resources and transactions used in this analytics example indicate the importance of a clear and accurate classification of transactions. Proper classification can affect the results allocation and the identification of an efficient level in the A/P process. The analysis in this pilot study uses the data as it was submitted, but the current review suggests the need to perform a clear classification process – standardization of transactions in future studies. The analysis results are highly sensitive to the transaction classification as will be shown in the next section. To illustrate this point different combinations of input and output were modelled.

5.2. Modelling

For our pilot investigation, using readily available data, the only inputs considered were the human resources, and the outputs were the most fundamental of A/P operations - the transactions processed. The volume of transactions is what makes this example "big data". For the inputs, we divided the human resource into five levels of officer, numbered 1-5. For the outputs, we identified three transaction types: invoices (I), journal vouchers (JV) and inter-departmental settlements (IDS). The latter two types may also be combined together as non-invoice types (NonI).

Fig. 2 shows the relative size of the different output types for all six organizations, though the axes have been scaled for confidentiality.

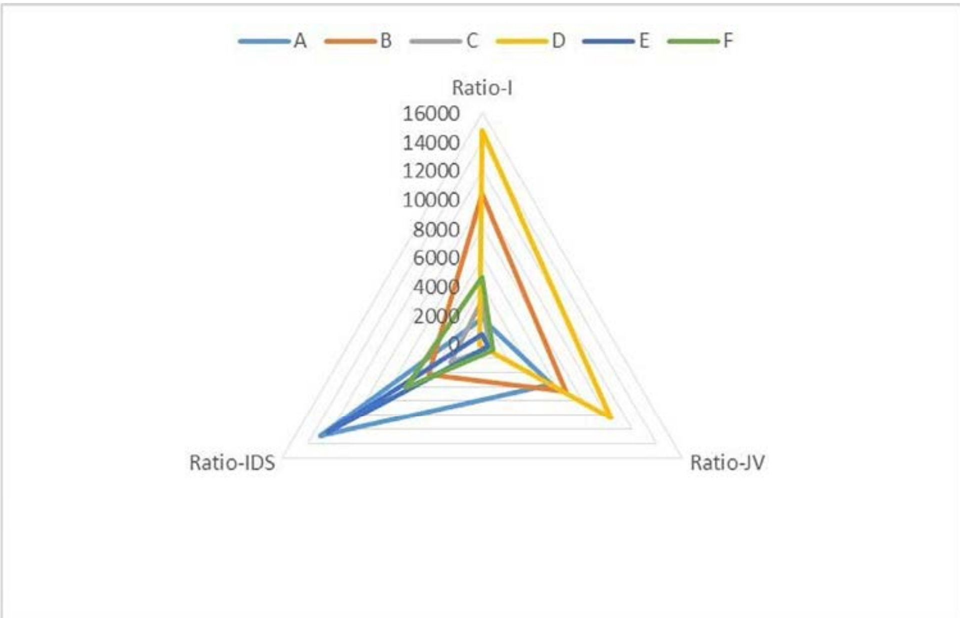

Fig. 2. Radar chart showing the different types of output for the six organisations.

Five different DEA models were investigated. Model 1 uses all five input types and all three output types. Model 2 uses all five input types, but just one output, total transactions (I+JV+IDS). Model 3 reverts to using all three output types, but uses only three input types and four of the organizations. This is because two of the organizations each use a (different) unique type of human resource. Model 4 is similar to model 3, but aggregates JV and IDS into NonI, thus giving only two output types. Finally, Model 5 uses all three output types and all six organizations, but aggregates all five input types based on total expenditure on human resources. This range of models is deliberately chosen to illustrate some of the important features that assumptions, knowledge and guidance need to bring to an analytics example. They are summarized in Table 1.

Table 1. Runs of the DEA models.

Model	Organizations (and number)	Input types (and number)	Output types (and number)
1	A to F (6)	1 to 5 (5)	I, JV, IDS (3)
2	A to F (6)	1 to 5 (5)	I+JV+IDS (1)
3	A, B, E, F (4)	2 to 4 (3)	I, JV, IDS (3)
4	A, B, E, F (4)	2 to 4 (3)	I, NonI (2)
5	A to F (6)	1+2+3+4+5 (1)	I, JV, IDS (3)

We will be relying on bar charts in our presentation of the results, for two reasons. First, as we mentioned above, the multi-input, multi-output nature of the models makes it difficult to visualize the results in two or even three dimensions. Second, one of the organizations is very much larger than the others: this distorts the scale of any graphs so much as to make them hard to display and read.

6. DEA results

6.1. Model 1

Fig. 3 shows the relative efficiency scores for model 1, which includes all organizations, all input types and all output types. This would be the obvious way for a novice data scientist, or anyone else inexperienced in analytics, to "do" DEA. The situation appears good, with all organizations except F apparently achieving a relative efficiency score of 100%. Yet the 100% scores for organizations C and D are particularly misleading, because each uses a unique input type. A feature of DEA is that such an organization will always appear to be 100% relatively efficient.

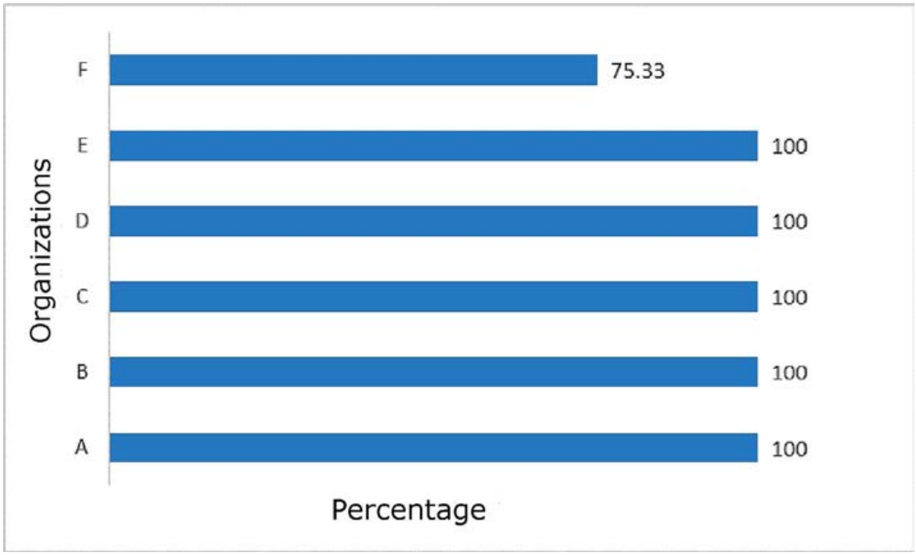

Fig. 3. Results for model 1 organizations, all input types, all output types.

6.2. Model 2

Model 2 combines the three transaction types into a single output. The relative efficiency scores are shown in Fig. 4. Organizations A and B still score 100% relative efficiency, as do (misleadingly, for the reason stated for model 1) organizations C and D. However, organizations E (15.51%) and F (33.93%) appear very inefficient. In DEA terms, 15.51% is a very low score, suggesting that the same outputs might be produced using less than one-fifth of the input resources.

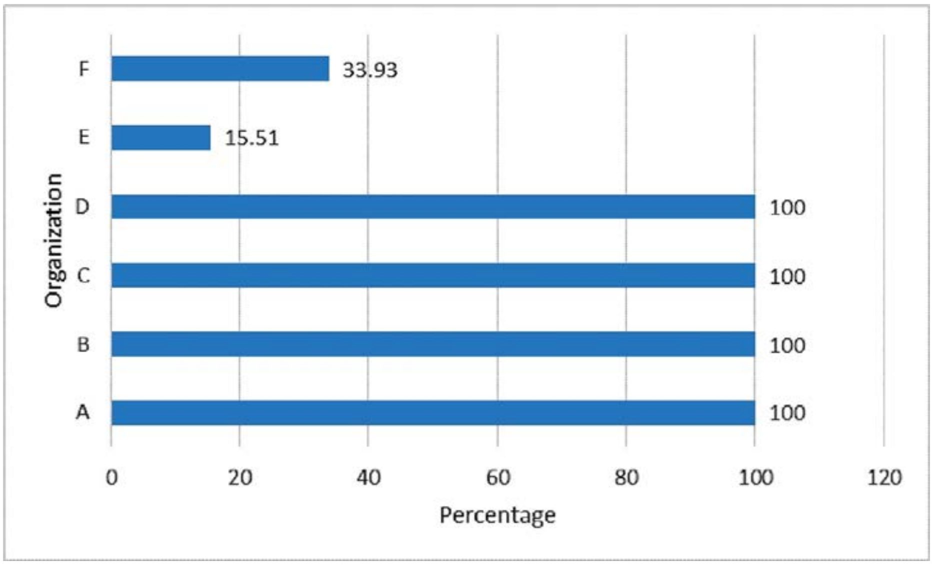

Fig. 4. Results for model 2: all organizations, all input types, single output.

6.3. Model 3

Excluding organizations C and D, and their associated human resource grades 1 and 5, gives a proper theoretically valid model 3. Its results are shown in Fig. 5. Organizations A and B remain at 100% relative efficiency, but are again joined by organization E. However, organization F is still relatively inefficient, at a score of 75.33%. These happen to be the same results as in model 1, but only by running this model could the analysts be sure that the efficiency scores for these four organizations were valid.

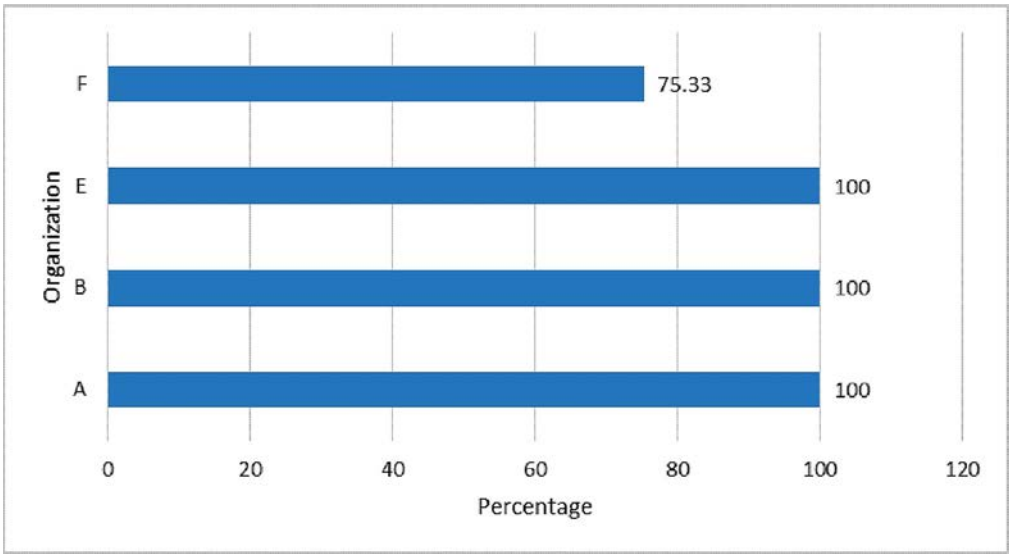

Fig. 5. Results for model 3: four organizations, three input types, all output types.

6.4. Model 4

Given the difference in results for models 2 and 3, where transactions were combined and separated respectively, the issue of transaction classification needs further investigation. Model 4 is thus the same as model 3, but with the output types JV and IDS combined to give NonI, so that there are only two output types. The results are shown in Fig. 6. Organizations A and B remain at 100% relative efficiency, and organization F is still relatively inefficient, at a score of 46.01%. Most interestingly, organization E is again very inefficient, with a score of 20.74%.

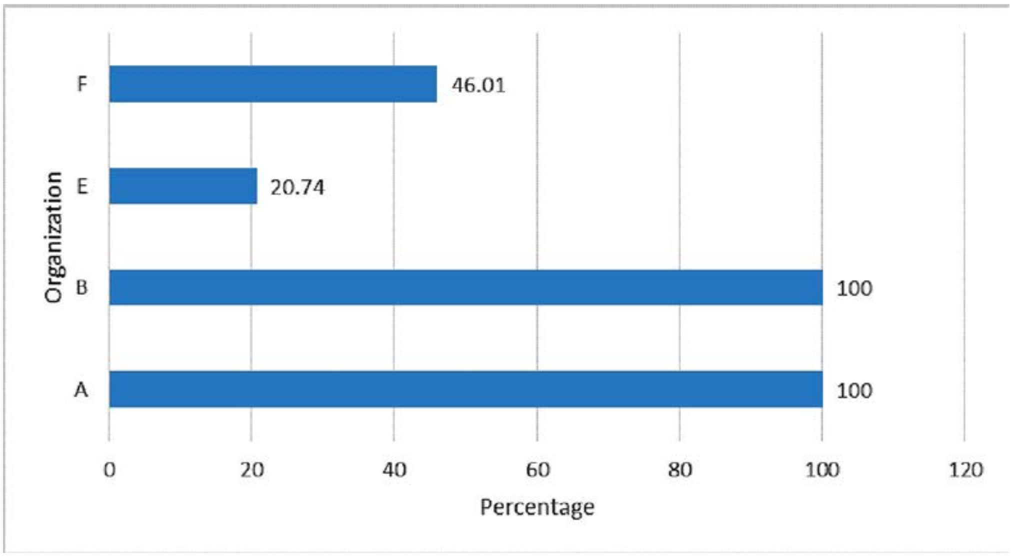

Fig. 6. Results for model 4: four organizations, three input types, two output types.

6.5. Model 5

In this model, all the inputs are combined into one type using salary expenditure. The results are shown in Fig. 7. Organizations A and D maintain their efficiency, with 100% relative efficiency scores, but organization C, freed from the misleading effect of its unique input, now has the lowest efficiency score, at 53.67%. Organization F scores 67.14%, E 90.93% and B 95.9%.

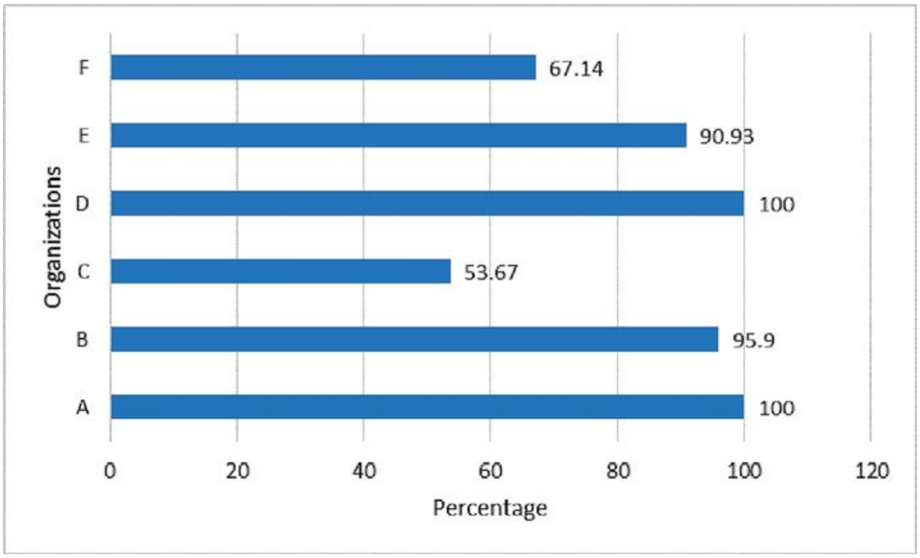

Fig. 7. Results for model 5: all organizations, all input types, single output.

7. Discussion

This pilot study shows some of the possibilities and limitations of analytics methods. Model 1 is the “just click the mouse” approach to using DEA on the A/P process data. This is what one might most likely get if DEA were available in an organization as “self-service analytics”. However, the results could be seriously misleading to anyone inexperienced in DEA; knowledge from both meaning and skilled practice comes into play here.

Model 2 is a simplified version, yet produces more meaningful and useful results than model 1, albeit only for four of the six of the organizations.

The results from models 3 and 4 illustrate that, even with the same output data, if it is classified in a different way, the apparent outcome of the DEA is different. This demonstrates how data and meaning interact, and shows why domain knowledge is so important when dealing with any data, big or otherwise. A data scientist with no domain knowledge would only be able to take the data as it comes from the system supplying it. As we see from these results, the outcomes will depend on the way in which that data is classified, and so any conclusions drawn may be misleading.

Model 5, by contrast, shows that just because an analytics model is mathematically possible and valid, it does not necessarily help to indicate any possible action. In this example, it is because the different types of human resource are not fully substitutable for each other. One cannot simply replace one senior member of staff with two very junior members, or *vice versa*, even if the salary levels are the same.

Even with these limitations, the results show that there exists a differentiation among the six organizations' efficiency level in the A/P process. In particular, organizations E and F, and possibly organization C, need to move to a more efficient level.

Further DEA studies may shed more light on these three, especially organization E, where the results of the different models vary considerably, and the other three organizations in the study. Other Canadian public sector organizations could be added to the study. Note of course that the inclusion of more organizations may change the results qualitatively, since DEA measures relative efficiency.

What DEA does not do, however, is prioritize the outputs. Alternative analytics techniques that can do that could be used as well as the DEA analysis, most obviously goal programming[17]. Since the assumptions of goal programming are different from those of DEA, knowledge of analytics would be crucial in a study that used both techniques; they are almost certain to produce somewhat different results.

The pilot study also relied on readily available data. A process-based study that collected data on the time taken and human resources used for each transaction would give a deeper understanding of the relationships, but automating the collection of such data might well face resistance from the staff concerned, particularly as this is the public sector.

This type of analytics study can identify multiple areas of improvement in process management. One of the most important things in managing strategic risk is through the benchmark review, to indicate ways to improve. To do this well, data consistency is crucial for the benchmarking exercise. There is important work to be done on data consistency by identifying what to include or not, even in something apparently as structured as the accounts payable process. For example, does the number of invoices include payroll, credit cards and so on? Equally, for DMUs to be compared, use of the same naming convention standards for the accounts and processes is essential. These actions are key to understanding, comparing and improving the same processes across organizations, or indeed across units within the same organization. Wrong fundamental inputs into the analysis reduce the capacity to control a risk that is inherent, storing up trouble for the future.

The purpose of this example is to show that despite the great potential of the DEA analytics technique for performance evaluation of a process, it will only be achieved with a knowledge-led application - like any other analytics technique.

8. Conclusions

8.1. From using analytics and big data

Analytics and big data have their usefulness now, as they have done under different names for the past 75 years or more. However, their application is nothing like as straightforward as articles in the computing press would have us believe. We have shown that it requires interaction between the data, the analytics techniques and knowledge - of both the techniques and the domain. The crucial need is for shared understanding, first of terminology (knowledge about data), then of interpretation (knowledge about meaning), then of action (knowledge about skilled practice). These come together as we have shown in Fig. 1. Indeed, in the longer term, over a series of analytics studies, there are three additional links, as shown in Fig. 8. One connects knowledge to data, for example, what data to collect, or indeed not bother to collect, and the issues about standardization (meaning) mentioned above. Another connects knowledge to techniques, developing new techniques in the light of past examples. The third connects action back to knowledge - learning from experience (skilled practice).

Some of the knowledge needed for that shared understanding can be structured and hence formalized - the explicit knowledge. For example, a DEA analytics package could flag when the results for a unit are unreliable because it uses a unique input type. However, the more tacit elements of the knowledge by definition cannot be structured in such a way, so human input will continue to be needed. The balance between the domain and technique knowledge required will vary from one study to the next. The amount of formalization possible will also vary, depending not just on the technique, but also on the expected users because of the basic need for shared understanding of terminology. The more heterogeneous the users, the less formalization will be possible.

To sum up, for a practical analytics study, first start with your purpose, then make sure you have enough data. Next, apply the tools of analytics, and understand the results. Crucially, then do something differently - and just as crucially, monitor what happens.

Most importantly, do not jump from one step to another without reflection. Big data, as we said above, can include a lot of garbage but probably the most difficult thing is the bias that in some cases the data can have. Social media is an example, when only a few very influential people are the participants in discussions. For instance, a "holding" organization can be affected in its management control system because of the lack of the important words integration and integrity.

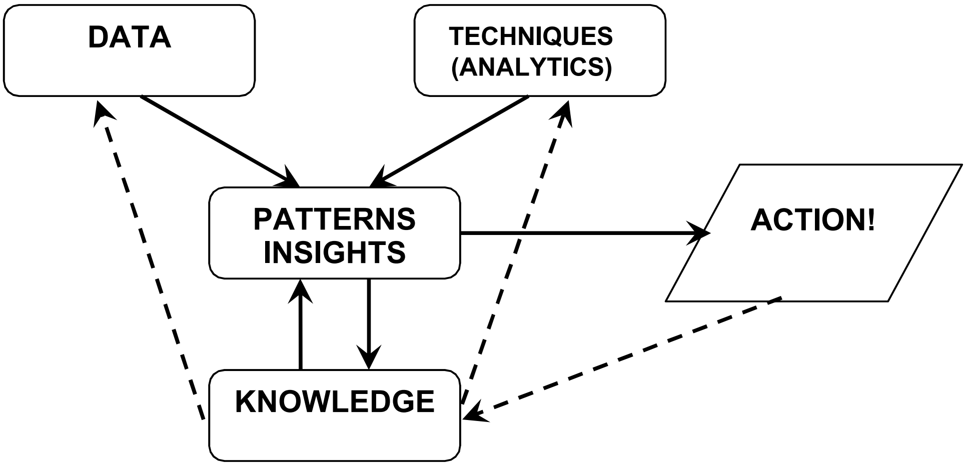

Fig. 8. The interactions between data, analytics and human knowledge over several studies.

8.2. From strategic risk perspective

This paper has presented the view of analytics and big data as a good opportunity for organizations to use intangible assets - data and analytics capabilities - that can be a source to protect the organization against multiple risks. However, the potential of these two assets can be jeopardized and morph into an extremely complex risk that is associated with invalid results. The value of filters, validation and review of all the analytics steps is more and more important, for several reasons. First, the risk of the creation of models to support decisions with invalid results, potentially leading to investment errors. This is common in budget selection when assignments are according to efficiency or regression parameters and the models are not correct. Second, the risk of creating confusion. Third, the risk of missing the focus of the problem that needs to be solved - doing what is easy rather than what is necessary. Finally, the risk of creating "wrong knowledge" or rather "increasing ignorance", or misusing assets.

References

1. Gates S. Incorporating strategic risk into enterprise risk management: A survey of current corporate practice. *Journal of Applied Corporate Finance* 2006; **18**: 81-90.
2. Young G, Hasler D. Managing reputational risks. *Strategic Finance* 2010; **92**:37-46.
3. Collins J, Ruefli T. Strategic risk: An ordinal approach. *Management Science* 1992; **38**:1707-1731.
4. Beer S. *Diagnosing the system for organisations*. Chichester: Wiley; 1985.

5. Davenport TH, Harris JG. *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business School Review Press; 2007.
6. https://www.informs.org/About-INFORMS/What-is-Operations-Research, accessed 22 April 2016
7. McAfee A, Brynjolfsson E. Big data: The management revolution. *Harvard Business Review* 2012; **90**: 60-68.
8. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 2012; **36**: 1165-1188.
9. Goes PB. Big data and IS research. *MIS Quarterly* 2014; **38**: iii-viii.
10. Kwon O, Lee N, Shin B. Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management* 2014; **34**: 387-394.
11. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics *International Journal of Information Management* 2015;
12. **35**: 137-144.
13. http://www.theguardian.com/world/2006/jul/20/secondworldwar.tvandradio, accessed May 5 2016
14. Spender JC. Thinking, talking, and doing: How the knowledge-based view might clarify technology's relationship with organizations.
15. *International Journal of Technology Management* 2007; **38**: 178-196.
16. Spender JC: Knowledge management: Origins, history, and development, In: Bolisani E, Handzic M (editors). *Advances in knowledge management: Celebrating twenty years of research and practice*. Switzerland: Springer International Publishing; 2015 p. 3-23.
17. Polanyi M. *The tacit dimension*. Garden City, NY: Doubleday; 1966.
18. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. *European Journal of Operational Research* 1978; **2**: 429-444.
19. Charnes A, Cooper WW. *Management models and industrial applications of linear programming*. New York: Wiley; 1961.